A RAND NOTE

AD-B168 760

DTIC
ELECTE
NOV 3 0 1992
S
C
D

Developing a Probability Sample of Prostitutes:
Sample Design for the RAND Study of HIV
Infection and Risk Behaviors in Prostitutes

Sandra H. Berry, Naihua Duan,
David E. Kanouse

July 1990

92-30359

RAND

92 11 30 015

A RAND NOTE

N-3190-NICHD

Developing a Probability Sample of Prostitutes:
Sample Design for the RAND Study of HIV
Infection and Risk Behaviors in Prostitutes

Sandra H. Berry, Naihua Duan,
David E. Kanouse

July 1990

RAND

Developing a Probability Sample of Prostitutes: Sample Design for the RAND Study of HIV Infection and Risk Behaviors in Prostitutes

Sandra H. Berry, Naihua Duan, and David E. Kanouse

Introduction

This paper outlines a preliminary sampling plan for a study of human immunodeficiency virus (HIV) infection and risk behaviors of Los Angeles prostitutes that will be carried out by The RAND Corporation. This study is now in the design phases. Pilot testing is scheduled for summer 1989 and fieldwork for fall 1989 through spring 1990.

Background

At present, acquired immune deficiency syndrome (AIDS) cannot be cured and no vaccine for preventing it has been developed. Consequently, efforts to contain the epidemic must emphasize changing behaviors that allow transmission of HIV, the AIDS-causing virus. The behavior of female prostitutes may significantly affect the epidemic's future, particularly its potential for spreading through heterosexual contact. Yet their behavior has been little studied and is poorly understood.

A study is being designed that will contribute to general understanding of heterosexual transmission by focusing on female prostitutes, their characteristics and behaviors, and the role they may play in the epidemiology of AIDS. Specific aims will include:
1. Developing numerical estimates of the size of the prostitute population in a large metropolitan area and of its distribution according to predominant mode of soliciting customers (street, out-call, massage parlor, escort service, brothel, etc.).
2. Characterizing prostitute career patterns

3. Performing HIV antibody testing to determine the extent of HIV infection in this population and how this varies by mode of solicitation.
4. Measuring the prevalence and incidence of specific risk behaviors (sexual and drug-related) that can transmit HIV infection.
5. Measuring the type and frequency of preventive behaviors (using condoms, disinfecting needles).
6. Examining the relationship between HIV antibody status, prostitute characteristics, and risk and preventive behaviors.
7. Estimating the numbers and percentages of specific sexual acts, both protected and unprotected, that occur between HIV-infected prostitutes and their customers, and the distribution of these acts according to prostitute characteristics.
8. Comparing the characteristics of the entire population of prostitutes with those of subgroups most likely to be recruited in studies of convenience samples (street prostitutes, prostitutes currently in jail, etc.).

The study will develop a statistical sampling frame and use it to identify, interview, and test a sample of 1000 prostitutes in Los Angeles County.

Collaborative research is now being carried out in various U.S. cities to determine how many prostitutes are infected with HIV-1. Virtually all the studies are using samples of convenience. Their results show seroprevalence rates from 0 percent (in Las Vegas, Nevada, and Colorado Springs, Colorado) to 57 percent (in Newark, New Jersey). Such studies provide valuable indications, but their statistical sampling techniques do not permit extrapolation to defined populations of epidemiologic interest. Instead, they provide information about selected groups of women who may differ substantially from those not sampled.

This study will provide unique data about prostitutes, permitting empirically based estimation of important population characteristics for the first time. By reducing uncertainty about this key population's characteristics and behavior, the study will greatly improve our ability

to construct epidemiologic models and predict the future course of the HIV-1 epidemic. It may suggest intervention strategies and ways to target them to the groups at highest risk of infection. Finally, it may improve our methods for collecting similar data in other geographic areas.

Overview of the Sampling Plan

Illegal markets are notoriously difficult to study because they are covert in nature. Fortunately (unlike gambling and drugs), the prostitution market depends heavily on advertising. Therefore, it may be possible to develop rough estimates of the number of prostitutes and to describe their market's characteristics. Further, previous studies suggest that despite their need to be secretive, many people in the business have cooperated when anonymity is guaranteed.

The central feature of this study design is the use of randomized sampling methods to produce unbiased estimates and to assess the estimates' precision, using standard statistical methods. Successful application of this sampling approach will have considerable research value, because it will dramatically improve estimates of the size and characteristics of the prostitute population.

Because no simple enumeration of all prostitutes is available, this population cannot be randomly sampled from a convenient list. Previous studies suggest that appropriate first-round sampling units are most easily constructed by stratifying prostitutes according to their means of soliciting clients.

We can distinguish five major solicitation media: (1) advertisements in mass media (newspapers and magazines); (2) listings in yellow pages (for massage parlors and escort services); (3) street signs (massage parlors, strip joints); (4) personal referrals (e.g., through bell captains, taxi drivers, and organizers of entertainment for events such as trade shows and conventions); and (5) personal solicitation (streetwalkers).

A combination of list sampling and area probability sampling will be used to construct an overall study sample including prostitutes who use each solicitation method. Based on previous research these can be grouped into three broad subpopulations: (1) streetwalkers; (2) sex industry workers; and (3) call girls.

Street Prostitutes

These women solicit primarily through physical presentation and are best studied using an area probability sample. The approach in this study will be one that has been used successfully in studies of the homeless—that is, using information in the first round from police and providers of social services to estimate the population density of prostitutes by block. These estimates will be checked and updated through field observations, and sampling will be employed to select blocks representing varying levels of nonzero density. Blocks will be sampled on a probability-proportional-to-size basis, where size equals density. During the second round, selected blocks will be sampled and attempts will be made to interview a specified sample of the prostitutes working there. The distribution of street prostitutes in an area changes by time of day and season of the year, so density estimates will have to be time-specific as well as area-specific.

Sex Industry Workers

Customers for sex industry workers are solicited in their places of employment (for example, massage parlors and clubs). To sample these work locations, a combination of list sampling and area probability sampling will be used, compiling lists from advertisements in yellow pages, newspapers, and magazines. High density areas will be located through informants, and will be identified through advertisements such as street signs. For each location, information about the number of women who work there and the percentage who are prostitutes will be gathered. Based on these estimates, locations for interviews will be selected and sample sizes defined at each location. Because gaining access to each location involves substantial fixed costs, the sampling will be allocated in clusters to reduce the number of locations.

Call Girls

This segment of the prostitute workforce may be the most difficult group to sample with traditional techniques, because they are more difficult to identify, count, and interview. This category includes women who work for or through escort services, as well as self-employed call girls. A list sampling approach will be used for this population, working with lists from various sources. First we will compile lists of advertised call girl services and their telephone numbers from the yellow pages and other published sources. These lists will be matched to eliminate duplicate telephone numbers. Then a random sample of telephone numbers will be drawn, again using a probability- proportional-to-size approach.

From lists of call girls obtained from taxicab drivers, bell captains, entertainment organizers, and others whose work puts them in a position to make such referrals, sample women will be contacted by telephone and a meeting arranged. Again, cluster sampling of women in services will be used to reduce costs.

Naturally, every effort will be made to minimize nonresponse, but some will occur; however, the sampling procedure will be adapted to minimize its effects. The strategy to be applied will ensure that persons or institutions that refuse to participate are replaced with others that are as similar as possible. Elements in each sampling frame will be stratified according to characteristics judged important and that can be measured in advance (for example, neighborhood characteristics, type of publication in which advertising appears). Refusals will be replaced with those elements in the sampling frame that are most similar with respect to a vector of such characteristics.

Overview of Data Collection

Size and Composition of the Prostitute Population. Information will be collected that will allow us to deter-

mine for each prostitute (1) all of the methods clients use to contact her (and thereby the sampling frames through which she might have been recruited), and (2) the periods when she was at risk of being recruited. When combined with information about the sampling probability (which is under our control) and the response rate (most of which is not), such information can be used to calculate the assumed sampling probabilities. When summed, the inverses of these rates yield estimates of the prostitute population's size.

Structured interviews will last between 60 and 90 minutes. In addition, a blood sample will be taken after appropriate counseling of each subject. Each subject will be paid $50 for participating in the study. Data will be collected anonymously, and subjects can get the results of their blood tests by collecting them in person at a specified date and time or by visiting an established testing and counseling center. Tests will be performed for HIV-1 antibodies, hepatitis B surface antibodies, human t-cell leukemia virus-1 and 2 antibodies, and syphilis. Follow-up counseling and results, if desired, will be available through the testing and counseling center or through referrals to social service agencies.

Although the procedures for data and blood sample collection will be the same for all sample strata, contact procedures will differ. For street prostitutes and sex industry workers, field staff, equipped with a van, will go out in teams. The teams will include a driver who doubles as a security guard (unarmed) and interviewers who are trained to draw blood samples and provide pretest counseling. The van will be used to store lab supplies and cash for respondent payments. It will provide a clean, well-lighted place for taking blood samples and for interviewing. If necessary, it will also provide a refuge and means of mobility in case of trouble.

Most prostitutes will be contacted on the job. If necessary, interviews will take place at a mutually convenient time outside working hours, to minimize their loss of income. However, many of the subjects will be interviewed during working hours; for them, the size of the payment may be important.

Because the data will be collected anonymously, the same person could be interviewed more than once. The field staff will be looking out for repeaters and will ask each prospective subject if she has been interviewed before. However, without some means of identifying the respondents or a staff large enough to cover an entire area in one 24-hour period, the possibility of some double counting cannot be eliminated.

Specific Problems Related to Sampling

Unit of Analysis

We plan to use a variety of units of analysis, including prostitutes, prostitute-client encounters, and person hours spent in prostitution. The choice of unit will depend on the purpose of the analysis. For example, in characterizing the prostitute population, the prostitute will be the unit of analysis. However, considerable variation in prostitutes' levels of activity is expected, with many prostitutes engaged in prostitution only on a part-

time basis. To characterize the prostitute work force, work force participation (person hours worked) will be used as the unit of analysis. To analyze the risk of HIV transmission, we will need to use the individual prostitute-client encounter as the unit of analysis.

For each unit chosen for reporting particular analyses, it is important that we be able to relate this unit to the unit of sampling, so that the sample can be weighted inversely proportional to the sampling probability. For the sample of street prostitutes, the sampling unit is approximated by person-hours spent on the street; a full-time prostitute has a higher probability of being included in the sample than a parttime prostitute. With this sampling approach, analyses that are based on person-hours as the unit of analysis are straightforward. For analyses that focus on prostitutes as the unit of analysis, a sample obtained cross-sectionally overrepresents full-time prostitutes; therefore, the sampled prostitutes must be weighted inversely by their level of work, defined by person-hours. For instance, a full-time prostitute in the sample should receive half the weight of a half-time prostitute in the sample, because the former is twice as likely to be sampled. To put it another way, if the characteristics of the population of all women engaging in prostitution are being estimated, the sample must be weighted to correct for the known overrepresentation of some types of prostitutes and underrepresentation of others. For analyses that are based on acts as the unit of analyses, the sample must be weighted by the encounter rates (number of encounters per unit of time worked).

For sex industry workers, the sampling unit will be the shops and the prostitutes who work at these shops. The list sample or area probability sample identifies the shops to be included in the sample; both shop-level and individual-level interviews will be conducted. To analyze using person-hours or encounters as the unit of analysis, the weights of the sample must be adjusted.

Nonresponse Bias

Reporting bias is a serious problem in human sexual research. In this study, many sampled prostitutes will be encountered who refuse to be interviewed, as well as respondents who are selectively cooperative; for example, who agree to be interviewed but decline to provide a blood sample. If the nonresponse is nonrandom, that is, if the respondents and the refusals differ in their HIV infection rate, the data would be affected by nonresponse bias: estimates based on the respondents would differ from what would have been obtained from the refusals.

Nonresponse bias is difficult to deal with in any survey research. We will mainly focus on evaluating the potential for having serious nonresponse, and hope that most of our major analyses do not have such a serious problem. If some of the analyses fail the test, the results will have to be qualified. However, the potential for nonresponse bias in this study should be substantially lower and the ability to evaluate that potential substantially higher than has been the case in prior studies based on volunteers or jailed prostitutes.

In many survey studies, the population studied is one that can be fairly well characterized by available data that are independent of the survey. In that case, to assess the potential for a serious nonresponse bias, it is possible to compare the characteristics of survey respondents with what is known about the population.

However, very little is known about the prostitute population that can be used for this type of analysis. Observable characteristics such as race, type of dress, approximate age, type of location, and so forth, of those who refuse to be interviewed can be collected. In addition, how respondents with similar observable characteristics to those who refuse compare with other types of respondents with respect to their reported behavior can be examined. This provides some information on the extent of response bias that is associated with these observable characteristics, but obviously provides none about any bias that is uncorrelated with these characteristics. This approach was used in a recent RAND study of the homeless, and will be made use of here as well.

Another approach is to use information on the difficulty of completing an interview with respondents as a way of judging possible differences between respondents and nonrespondents. Some interviews are more difficult to obtain than others; for example, they require more interviewer persistence or persuasion. Reluctant and difficult-to-reach respondents may offer clues to the char-

acteristics of nonrespondents, who are even more reluctant or difficult to reach. If measures are taken of the difficulty of obtaining each interview, it is then possible to examine, within the sample of completed interviews, the relationship between completion difficulty and respondent characteristics on the one hand and responses to key items on the other. This provides one basis for assessing possible biases introduced by nonresponse. Although hardly a definitive solution to the problem, this is feasible and worth doing.

Still another approach is to offer additional incentive payments to a random subsample of refusals, and then to compare the responses of those initially refusing with those of other respondents to gain some idea about the distinctiveness of nonrespondents. This was considered as a possible strategy, but rejected as infeasible in a field study of this population, where an active grapevine can be expected to quickly broadcast news of any differential incentives.

For these reasons, the nonresponse bias will be dealt with (1) by taking all feasible measures to minimize the extent of nonresponse, (2) gathering as much information as possible on the characteristics of nonrespondents, (3) measuring interview completion difficulty for respondents, and (4) analyzing data gathered in (2) and (3) to assess how nonrespondents might differ from respondents in their characteristics and behavior.